

University of Groningen

On the color of voices

El Boghdady, Nawal

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

El Boghdady, N. (2019). *On the color of voices: the relationship between cochlear implant users' voice cue perception and speech intelligibility in cocktail-party scenarios*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



English Summary



1. BACKGROUND

Cochlear implants (CIs) are surgically-implanted neuroprosthetic electrical devices that restore the sensation of hearing in deaf and hard-of-hearing individuals. In a CI, signal processing algorithms quantize incoming sounds into several discrete frequency channels which are then converted into electrical signals to stimulate the designated implanted electrodes. While CI users' speech intelligibility scores in quiet are usually good, understanding speech masked by background noise, especially another competing talker, is still quite challenging. In a survey of 247 adult CI users, while 93% of the respondents indicated their satisfaction with speech intelligibility in quiet, only about 30% indicated a similar experience for speech perception in noise. Speech masked by a competing talker is thus considered an important problem for CI users because it describes everyday listening situations, such as family gatherings, outings in crowded areas, and work environments consisting of open-space offices.

To differentiate between speakers, normal hearing (NH) listeners utilize several acoustic cues including the speaker's pitch, manner of articulation, regional accent, and speaking style. Together, these cues contribute to the color of the speaker's voice. This dissertation mainly focused on two fundamental voice cues that arise from the speaker's anatomy, namely, the voice pitch (F0) and the vocal tract length (VTL), as they have been shown consistently in the literature to be not only related to the perception of the speaker's age, gender, and identity in NH listeners, but also related to speech intelligibility in the presence of a competing talker (speech-on-speech; SoS). The literature has also shown that CI users experience difficulties perceiving those two cues, with VTL cue perception constituting a higher challenge than F0 cue perception. Thus, a key aim of this dissertation was to investigate whether the difficulties experienced by CI users in SoS perception were related to their

inability to properly perceive F0 and VTL cues, and whether optimizing implant parameters, such as signal processing algorithms, stimulation patterns, or the frequency-to-electrode allocation mapping, could help improve the perception of such cues in CI listeners.

2. METHODOLOGICAL APPROACH

To address these research questions, in this dissertation, I developed experimental paradigms that were based on existing psychoacoustic tasks and speech material from the literature. In the first task, I measured SoS intelligibility, the per-word intelligibility of the target speaker as it was masked by a concurrent talker, as a function of increasing the F0 and VTL difference between the two simultaneous talkers. F0 and VTL of the original target speaker were synthetically manipulated to produce the desired masker speaker's voice. In the second task, I measured SoS comprehension as a function of the F0 and VTL difference between the target and masker speakers in a similar way to that applied in the first task. For SoS comprehension, participants were requested to assess whether the entire target sentence made sense or not, as opposed to SoS intelligibility in which participants were scored based on the number of words correctly repeated from the target sentence. In the third task, I measured participants' sensitivities to F0 and VTL by computing the smallest F0 and VTL differences they could reliably detect; the just-noticeable-difference (JND).

I tested different voice spaces for F0 and VTL differences, namely, the child speaker space and the male speaker space relative to the female speaker space. This was driven by the expectation that these voice spaces may be represented differently in the implant stimulation pattern, and hence, may yield different speech perception results in CI users. Additionally, to obtain a better generalizability of the results, I ran my experiments on a range of CI users spanning more than one native language:

a Dutch-speaking group (Chapters 2 and 5) and a German-speaking group (Chapters 3 and 4). This involved porting the Dutch tasks to the German language.

3. RESEARCH QUESTIONS AND FINDINGS OF THIS DISSERTATION

The three psychoacoustic tasks were deployed to answer four main questions, to each of which a separate chapter was devoted. In Chapter 2, the question of whether SoS perception (intelligibility and comprehension) is related to CI users' sensitivity to F0 and VTL differences was investigated. SoS intelligibility and comprehension in the female-child speaker space were assessed in both Dutch-speaking NH and CI listeners, and F0/VTL JNDs were measured for the CI group. The data revealed that while NH listeners gained a benefit in SoS perception from increasing the F0 and VTL differences between a female target speaker and a child masker, CI users on the contrary did not and, in fact, demonstrated a small but significant decrement in performance with an increasing VTL difference. This decrement was found to be correlated with the CI users' sensitivity to VTL differences. Additionally, CI users' overall SoS perception was correlated with both F0 and VTL sensitivity, but not to either of them in isolation. These findings highlight the importance of voice cue perception in SoS-related tasks.

In Chapter 3, the question of whether parasitic channel interaction in the implant could influence the relationship between SoS perception and voice cue sensitivity in CI users was investigated. Channel interaction in the implant is a side-effect of electrical stimulation inside the conductive fluid-filled cochlea. Current spreads across neighboring electrodes introducing cross-talk and, subsequently, reduces the spectral resolution in the implant. SoS perception was assessed in the region encompassing female to male voice transitions in the F0-

VTL space. The data revealed that increased channel interaction could play an important role in the degradation of both SoS perception and reduced F0/VTL sensitivity in CI users. In addition, F0 and VTL manipulations of female voices towards male-like voices yielded an improvement in SoS perception in CI users. This finding contrasted with the observation in Chapter 2 regarding manipulations from female voices towards child-like masker voices.

In Chapter 4, the potential of improving the relationship between SoS perception and voice cue sensitivity in CI users using a spectral contrast enhancement (SCE) algorithm was assessed. The SCE algorithm attempts to enhance the representation of the spectral peaks in the signal as these peaks carry important formant frequency information and may encode essential VTL-related cues. Results revealed that while SCE had the potential of improving SoS intelligibility (there was no observable effect on SoS comprehension), sensitivity to either F0 or VTL differences remained unaffected. This outcome indicates that SCE could effectively be improving the overall signal-to-noise ratio of the SoS input rather than enhancing the sensitivity to the underlying voice differences between the two competing talkers.

Finally, in Chapter 5, the potential of manipulating the frequency-to-electrode allocation map to improve VTL sensitivity was investigated in vocoder simulations of CI processing with NH listeners. The frequency-to-electrode allocation map in a CI defines how the acoustic frequency range of the incoming signal is to be quantized and distributed among the implanted electrodes. The data revealed that when an insufficient number of frequency channels (simulated electrodes) is assigned to the range below 3 kHz, where most formant frequency information is present, sensitivity to VTL differences is lowest. When the number of such channels increases, sensitivity to VTL differences improves. These findings indicate that the careful optimization

of frequency-to-electrode allocation maps may be crucial for successfully transmitting important VTL-related information. In the clinic, such maps are seldom customized for each participant and because of the variability in cochlear dimensions, electrode array lengths, and insertion depths, the standard frequency-to-electrode map is rarely aligned with the corresponding place of stimulation along the basilar membrane.

4. CONCLUSIONS

CI users experience difficulties in SoS perception that appear to be related to their low sensitivity to F0 and VTL differences. These difficulties could be enhanced by applying advanced signal processing algorithms or channel stimulation patterns optimized for reducing parasitic channel interaction. Another approach that may be followed either in isolation or in combination with the aforementioned solutions would be to optimize implant parameters such as the frequency-to-electrode allocation map. Such approaches should be carried out bearing in mind that the representation of voice cues in the implant vary depending on the nature of the voice space (male, female, or children's voice spaces). This indicates that an optimization technique may enhance the representation, and hence perception, of a group of voices at the expense of the others. The conclusions from this dissertation underscore the importance of optimizing implant parameters on an individual basis and that a one-size-fits all optimization strategy would not yield optimum results in the CI population as a whole.